Acknowledgements
We thank D. Tard and M. Montellh-Zoller for technical assistance. This work was funded in part by a Beth Israel Pathology Foundation grant, a BIDMC Emerson Fellowship award to A.M.S. and NIH grants to A.M.S. and M.J.B. A.L.P. is supported by the German Academic Exchange Service (DAAD). In memory of Q. Zhu.

Correspondence and requests for materials should be addressed to A.M.S. (e-mail: andrewms@u.washington.edu). The GenBank accession number for NUDT9 cDNA and protein sequences is AY026252.

Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

Jean-Pierre Hugot*†‡, Mathias Chamaillard*†, Habib Zouali*, Suzanne Lesage*, Jean-Pierre Cézard‡, Jacques Belaiche§, Sven Almer‖, Curt Tysk¶, Colm A. O'Morain#, Miquel Gassull☆, Vibeke Binder**, Yigael Finkel††, Antoine Cortot‡‡, Robert Modigliani§§, Pierre Laurent-Puig†, Corine Gower-Rousseau‡‡, Jeanne Macry‖‖, Jean-Frédéric Colombel‡‡, Mourad Sahbatou* & Gilles Thomas*†¶¶

* *Fondation Jean Dausset CEPH, 27 rue J. Dodu 75010 Paris, France*
† *INSERM U434, 27 rue J. Dodu 75010 Paris, France*
‡ *PEWG-IBD, Department of Paediatric Gastroenterology, Hôpital Robert Debré, 75019 Paris, France*
§ *GETAID and Department of Gastroenterology, CHU de Liège, 4000 Belgium*
‖ *Division of Gastroenterology and Hepatology, IHM, Linköpings Universitet, SE-581 85 Linköping, Sweden*
¶ *Department of Gastroenterology, Örebro Medical Center Hospital, SE-701 85 Örebro, Sweden*
\# *Department of Gastroenterology, Adelaide & Meath Hospital, Dublin 24, Ireland*
☆ *Department of Gastroenterology, Hospital Universitari Germans Trias i Pujol, Badalona, 08916 Spain*
** *Department of Gastroenterology Herlev Hospital, 2730 Herlev, Denmark*
†† *Astrid Lindgren Children's Hospital, SE-161 76 Stockholm, Sweden*
‡‡ *Registre EPIMAD, Hôpital Calmette, 59037 Lille, France*
§§ *Department of Gastroenterology, Hopital Saint Louis, 75475 Paris, France*
‖‖ *INSERM U458, Hôpital Robert Debré, 75019 Paris, France*
¶¶ *Department of Surgery, Hôpital Saint Antoine, 75012 Paris, France*

Crohn's disease[1,2] and ulcerative colitis, the two main types of chronic inflammatory bowel disease, are multifactorial conditions of unknown aetiology. A susceptibility locus for Crohn's disease has been mapped[3] to chromosome 16. Here we have used a positional-cloning strategy, based on linkage analysis followed by linkage disequilibrium mapping, to identify three independent associations for Crohn's disease: a frameshift variant and two missense variants of *NOD2*, encoding a member of the Apaf-1/Ced-4 superfamily of apoptosis regulators that is expressed in monocytes. These *NOD2* variants alter the structure of either the leucine-rich repeat domain of the protein or the adjacent region. NOD2 activates nuclear factor NF-kB; this activating function is regulated by the carboxy-terminal leucine-rich repeat domain, which has an inhibitory role and also acts as an intracellular receptor for components of microbial pathogens. These observations suggest that the *NOD2* gene product confers susceptibility to Crohn's disease by altering the recognition of these components and/or by over-activating NF-kB in monocytes, thus documenting a molecular model for the pathogenic mechanism of Crohn's disease that can now be further investigated.

Crohn's disease (CD; MIM 266600) occurs primarily in young adults with an estimated prevalence of 1 in 1,000 in western countries[2]. Its incidence has increased markedly over the past half century, arguing for the involvement of recent, unidentified, environmental factors[2]. Familial aggregation of the disease suggests that genetic factors may also be involved—an hypothesis that was substantiated in 1996 by the discovery of a susceptibility locus for CD, *IBD1*, on chromosome 16 (ref. 3). Identification of the exact nature of the genetic changes that are implicated in CD susceptibility would provide a specific approach to understanding this common disorder.

Because candidate genes previously localized on chromosome 16 failed to show an association with CD[4,5], we refined the localization of the *IBD1* susceptibility locus by typing 26 microsatellite markers spaced at an average distance of 1 cM in the pericentromeric region

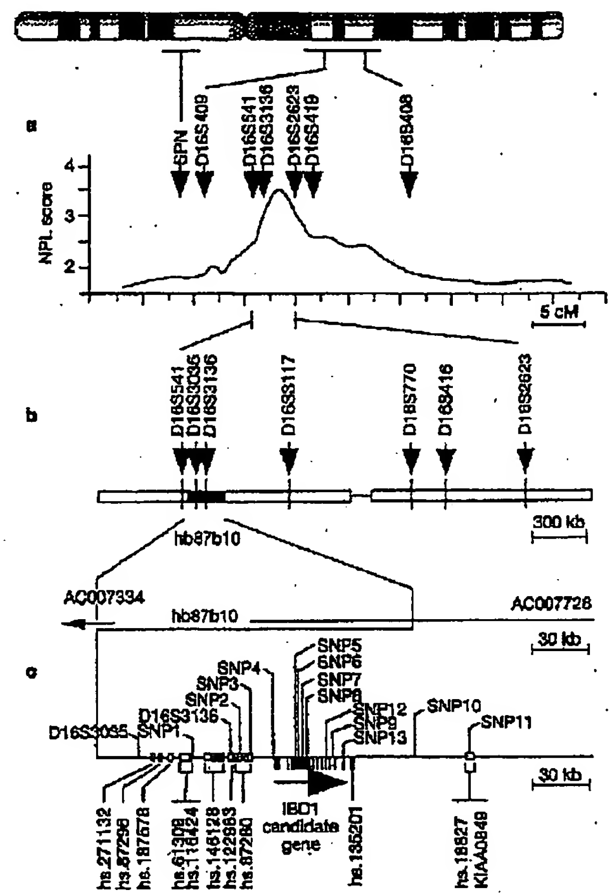

Figure 1 Strategy used to identify the *IBD1* locus. **a**, Profile for the multipoint nonparametric linkage (NPL) scores. Approximate cytogenetic localizations are shown for selected microsatellite markers used in the study[3] that localized *IBD1* to the pericentromeric region of chromosome 16. Subsequent linkage analyses were focused on the region between SPN and D16S408. The highest NPL score (maximum NPL score, 3.49; $P = 2.37 \times 10^{-4}$) was in the region between markers D16S541 and D16S2623 (ref. 6). **b**, Physical map of the *IBD1* region[6]. White and black boxes correspond to the two BAC contigs and BAC clone hb87b10, respectively. Five yeast artificial chromosomes (YACs) bridge a gap of ~100 kb between these two contigs. The position on the physical map of the microsatellite markers used in the linkage analysis is indicated. Distance between D16S541 and D16S2623 is ~2 Mb. **c**, Representation of the sequenced region containing the *IBD1* candidate gene. Unigene clusters and 11 exons of the *IBD1* candidate gene are indicated by white and black boxes, respectively. Bold horizontal arrow indicates direction of transcription. Positions of SNP 1–13, D16S3035 and D16S3136, which were typed in 235 CD families for linkage disequilibrium studies, are shown.

of chromosome 16. Model-free linkage analyses performed on 77 multiplex families indicated that the probability was 0.7 for the location of the susceptibility locus between D16S541 and D16S2623 (Fig. 1a). We constructed bacterial artificial chromosome (BAC) contigs spanning this region (Fig. 1b), which supported linkage disequilibrium mapping[4]. The transmission disequilibrium test[5], performed on a single trio from each of 108 (77 multiplex and 31 simplex) families, showed a borderline significant association ($P < 0.05$) between the disease phenotype and the 207-base-pair (bp) allele of D16S3136. This observation was replicated with another set of 76 families, although with a different allele (the 205-bp allele; $P < 0.01$). These two observations may be due to type-one errors. Alternatively, they may reflect true association in two sets of families drawn from genetically different populations.

The latter hypothesis led to the following strategy: a 164-kb BAC clone (hb87b10) from the CEPH-BAC library containing D16S3136 was sequenced completely (EMBL accession number AJ303140). A public database search extended the sequence of the corresponding region to 260 kb but did not identify characterized genes, with the exception of KIAA0849, which codes for a ubiquitin C-terminal hydrolase homologue in *Caenorhabditis elegans*. However, analysis by GRAIL and an expressed sequence tag (EST) homology search identified many putatively transcribed regions (Fig. 1c).

Eleven single-nucleotide polymorphisms (SNP 1–11) selected from these regions were genotyped, together with microsatellite markers D16S3035 and D16S3136, in a total of 235 available CD families (Table 1). Strong linkage disequilibrium was observed among most markers (data not shown). Several SNPs showed significant association with CD by the pedigree disequilibrium test[6] (PDT), confirming the existence of linkage disequilibrium, with the disease locus over the investigated region (specially SNP 2, nominal P value 0.00002; Table 1).

These observations prompted the characterization of neighbouring Unigene clusters (Fig. 1c). Eleven overlapping clones, isolated from a human leukocyte complementary DNA library, extended Unigene cluster hs135201 and identified 11 exons of a single gene. The previously identified SNP 5–8 were contained in exon 3 of this gene and shown to be non-synonymous variants. To find additional disease-related variants, all exons of this gene were sequenced in 50 unrelated CD patients—each a member of an affected sibling pair identical by descent for both chromosome 16 homologous regions. Two additional non-synonymous SNPs (SNP 12 and 13), with rare-allele frequencies greater than 0.03, were identified and subsequently used to type the 235 CD families.

The PDT was most significant for SNP 13 ($P = 6 \times 10^{-5}$). Families were divided into two groups: those with at least one member carrying the rare allele of SNP 13 and those without this allele. The latter group of families failed to show association between CD and SNP 4–6, and showed considerable decrease in the significance of the SNP 2 association. This result indicates that the associations of these four loci with CD were not independent of SNP 13 (Table 1). In contrast, significance of the CD associations with SNP 8 and 12 decreased modestly in these families, indicating a minimal contribution of the rare SNP 13 allele to these associations.

The 8 intragenic SNPs that were initially identified defined 41 different haplotypes. Three of these revealed preferential transmission to affected individuals (Table 2). These three haplotypes each contain one rare allele of SNP 8, 12 or 13 in a context of a common background. Notably, the haplotype defined by the same background and by the absence of these rare alleles did not show such transmission distortion (Table 2). Furthermore, the rare alleles of SNP 8, 12 and 13 were never found on the same haplotype, indicating independent association of CD susceptibility with each of three non-synonymous variants of a same gene.

As a result of these associations, the allele frequencies of SNP 8, 12 and 13 differed in the group of CD patients as compared with controls (Table 3). Average risks for CD, computed for genotypes containing zero, one or two variants (Table 3), revealed a gene-

Table 1 Linkage disequilibrium analyses at the *IBD1* locus

Marker name	Distance to next marker (bp)	Rare allele frequency in CD family founders*	Nominal *P*-values for the PDT in CD families†			Marker location/structural features‡
			All (N = 235)	SNP 13+ (N = 65)	SNP 13– (N = 170)	
D16S3035	29,389	Microsatellite marker	NS	0.005	NS	
SNP1	19,434	0.34	NS	NS	NS	End of BAC hb133D1
D16S3136	3,273	Microsatelite marker	NS	0.002	NS	Exon of hs.122983
SNP2	5,459	0.25	0.00002	0.00008	0.02	Exon of hs.87280
SNP3	15,076	0.35	NS	0.005	NS	Exon of hs.87280
SNP4	14,398	0.40	0.0008	0.000003	NS	Grail putative exon
SNP5	576	0.39	0.0001	0.00001	NS	*IBD1* exon 3 721C>T P241S
SNP6	385	0.38	0.0001	0.000006	NS	*IBD1* exon 3 1295C>T R432R
SNP7	344	0.35	NS	0.0003	NS	*IBD1* exon 3 1890T>T R432R
SNP8	10,593	0.10	0.001	0.03	0.009	*IBD1* exon 3 2023C>T R676W
SNP12	2,727	0.03	0.003	NS	0.0041	*IBD1* exon 7 2641G>C G1861R
SNP9	4,510	0.37	0.01	0.00003	NS	*IBD1* intron 8 IVS8+133delAinsCT
SNP13	35,121	0.07	0.000006	0.000006	–	*IBD1* exon 10 2936insC 980fs981X (frame shift)
SNP10	28,592	0.37	0.05	0.002	NS	End of BAC hb27G11
SNP11		0.09	NS	NS	NS	Exon 9 of KIAA0849 1058G>T D352D

*Markers are ordered from centromere to telomere. The nomenclature used does not take into account an alternative NOD2 exon located upstream of IBD1 exon 1 (ref. 14). NS, not significant.
†Values computed for the 599 CD family founders.
‡Values for pedigree disequilibrium test (PDT) are reported for all families and the subgroups of families with or without the rare allele of SNP 13.
§When relevant, its effect on the coding sequence are indicated.

dosage effect. The major increase in risk associated with two variant alleles confirms the recessive nature of CD susceptibility, which was suggested by previous segregation analysis[9–11] and linkage studies[1]. It may also contribute to explain the unusual precision of the affected sibling-pair analysis in mapping the susceptibility gene.

The rare allele of SNP 8 was associated positively with the 205-bp allele of D16S3136 and negatively with the 207-bp allele. The inverse association was noted for the rare alleles of SNP 12 and 13, thus providing a rationale for the initial observations made with this microsatellite marker (data not shown). Genotype frequencies were comparable in CD patients originating from uniquely and multiply affected kindred—an observation compatible with the close clinical similarity of the sporadic and familial diseases[12]. The observed linkage of CD to chromosome 16 could not be entirely explained by the present associations, because GeneHunter analysis of 85 multiplex families without SNP 8, 12 and 13 revealed a component of linkage (nonparametric lod score (NPL) 1.6, pointwise significance $P < 0.02$). Thus, other variants of this gene or additional genes on chromosome 16 may be involved in CD susceptibility.

Genotyping of 167 patients with ulcerative colitis revealed genotype frequencies comparable to those of controls, indicating that these SNPs were not associated with susceptibility to ulcerative colitis—an observation in agreement with its lack of linkage to the *IBD1* locus[13].

The candidate *IBD1* gene has high expression in leukocytes, but low or no expression in the other investigated tissues, including normal colon and small intestine (results not shown). It encodes a 1,013-amino-acid protein that is identical to NOD2—a member of the CED4/APAF1 superfamily of apoptosis regulators[14]. From its amino terminus to its carboxy terminus, NOD2 is composed of two caspase recruitment domains (CARD), a nucleotide-binding domain (NBD), and a LRR region (Fig. 2b). The LRR domain of NOD2 has binding activity for bacterial lipopolysaccharides[15] (LPS) and its deletion stimulates the NF-kB pathway[16–18].

The rare allele of SNP 13 corresponds to a 1-bp insertion in exon 10 (980fs) predicted to truncate NOD2 in the LRR region. Those of SNP 8 and 12 cause non-conservative substitutions in the LRR domain (G881R) and in the proximal adjacent region (R675W), respectively (Fig. 2a). Systematic sequencing of the coding sequence of NOD2 revealed additional very rare missense variants, which together were observed in 5% of controls and 4% of patients with ulcerative colitis. This percentage rose to 17% for CD patients, where the most frequent variants tended to cluster in the LRR and its adjacent regions (Fig. 2b). This excess suggests that, in addition to SNP 8, 12 and 13, more variants in this part of the NOD2 protein may be associated with CD susceptibility. Thus, the LRR domain of CD-associated variants is likely to be impaired, possibly to various degrees, in its recognition of microbial components and/or in the physiological inhibition of NOD2 dimerization, thus resulting in the inappropriate activation of NF-kB in monocytes.

Much evidence supports bacteria-induced NF-kB disregulation in CD. First, susceptibility to spontaneous inflammatory bowel

Table 2 Transmission disequilibrium of *NOD2* haplotypes in CD families

SNP no.	Haplotype								Transmitted	Non-transmitted	PDT (P-value)
	4	5	6	7	8	12	9	13			
SNP 8	1	1	1	1	2	1	1	1	4	1	
	2	1	1	1	2	1	1	1	2	0	
	2	**1**	**1**	**1**	**2**	**1**	**2**	**1**	82	48	0.001
	1	1	1	1	2	1	2	1	3	4	
	2	1	2	1	2	1	2	1	7	0	
	2	2	2	1	2	1	2	1	0	1	
SNP 12	1	2	2	2	1	2	1	1	1	3	
	2	1	2	1	1	2	1	1	0	1	
	2	**1**	**1**	**1**	**1**	**2**	**2**	**1**	38	13	0.0005
	1	1	1	1	1	2	2	1	3	2	
	2	2	1	1	1	2	2	1	1	1	
	2	2	2	1	1	2	2	1	2	2	
	2	2	2	2	1	2	2	1	3	1	
	1	2	2	2	1	2	2	1	1	0	
	1	1	1	2	1	2	2	1	1	0	
SNP 13	2	1	1	1	1	1	1	2	2	1	
	1	1	1	1	1	1	1	2	0	1	
	2	**1**	**1**	**1**	**1**	**1**	**2**	**2**	83	22	0.0000002
	1	1	1	1	1	1	2	2	2	0	
	1	2	2	1	1	1	2	2	1	0	
	2	2	2	2	1	1	2	2	0	1	
None*	1	2	2	2	1	1	1	1	116	141	NS
	2	2	2	2	1	1	1	1	2	4	
	1	1	2	2	1	1	1	1	0	0	
	1	2	1	2	1	1	1	1	0	2	
	1	1	1	2	1	1	1	1	1	0	
	2	1	1	2	1	1	1	1	1	1	
	1	2	2	1	1	1	1	1	9	19	NS
	1	1	1	1	1	1	1	1	4	7	
	2	1	1	1	1	1	1	1	0	0	
	2	2	2	1	1	1	1	1	2	0	
	2	2	2	2	1	1	2	1	7	4	
	1	2	2	2	1	1	2	1	9	7	
	2	1	2	1	1	1	2	1	0	1	
	1	1	2	1	1	1	2	1	1	1	
	1	2	2	1	1	1	2	1	94	116	NS
	2	2	2	1	1	1	2	1	20	16	NS
	2	**1**	**1**	**1**	**1**	**1**	**2**	**1**	70	78	NS
	1	1	1	1	1	1	2	1	11	7	
	2	2	1	1	1	1	2	1	0	1	
	1	2	1	1	1	1	2	1	3	9	

All haplotypes defined by typing 8 intragenic SNPs of *IBD1* in 235 CD families are indicated. They are classified according to the presence of a rare allele of SNP 8, SNP 12 or SNP 13. The number of transmitted or non-transmitted haplotypes from heterozygous parents to their affected offspring is shown. The statistic used to evaluate transmission distortion is PDT[6]. Haplotypes shown in bold are discussed in the text.
*Includes all haplotypes where none of the rare alleles of SNP8, 12 or 13 are present.

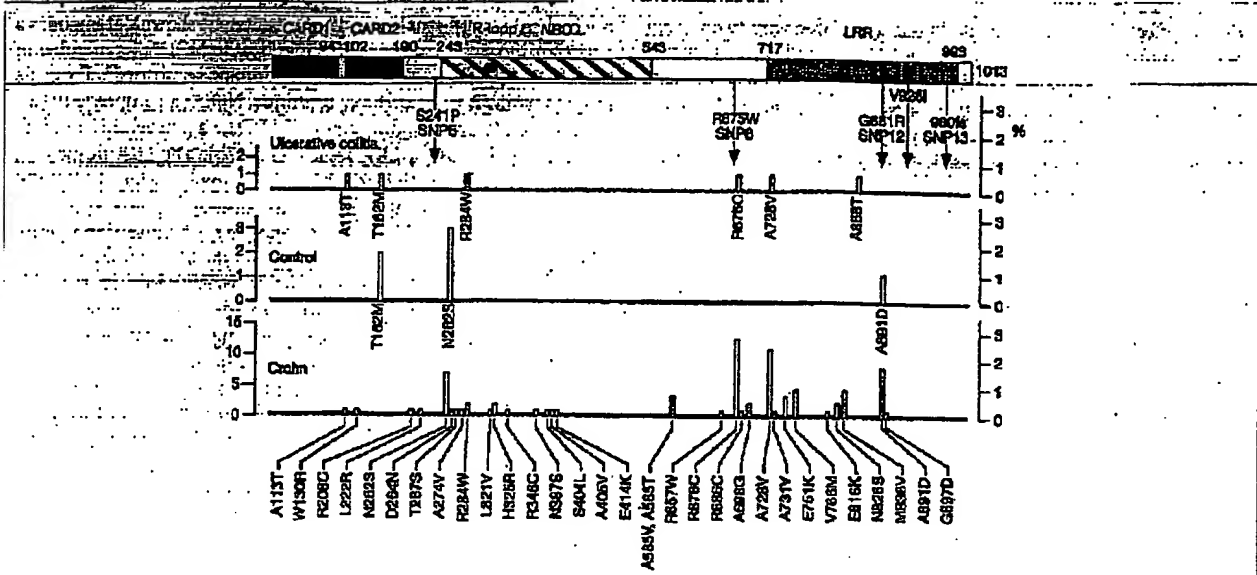

Figure 2 Representation of the IBD1/NOD2 protein variants. The translation product deduced from the cDNA sequence of the candidate *IBD1* gene is identical to that of *NOD2* (ref. 14). The polypeptide contains two caspase recruitment domains (CARD), a nucleotide-binding domain (NBD) and ten 27-amino-acid, leucine-rich repeats (LRRs). Black circle indicates the consensus sequence of the ATP/GTP-binding site motif A (P-loop) of the NBD. The sequence changes encoded by the three main variants associated with CD are SNP 8 (R675W), SNP 12 (G881R) and SNP 13 (980 frameshift). This frameshift changes a leucine to a proline at position 980, and is immediately followed by a stop codon. SNP 5 is described in Table 1. The allele frequencies of the V928I polymorphism were not significantly different (0.92±0.08) in the three groups, and the corresponding genotypes were in Hardy–Weinberg equilibrium. The positions of the rarer missense variants, observed in 457 CD patients, 159 ulcerative colitis patients and 103 unaffected unrelated individuals, are indicated for these groups. Left scale indicates the number of each identified variant in the investigated groups; right scale measures the mutation frequency.

disease (IBD) in mice has been associated with mutations in Toll-like receptor 4 (TLR4)—a member of a family of NF-κB activators that is known to bind LPS through its LRR domain[19,20]. Second, antibiotic therapy causes transient improvement of CD patients, supporting the hypothesis that enteric bacteria may have an aetiological role in CD[21]. Third, NF-κB has a pivotal role in IBD and is activated in mononuclear cells of the intestinal lamina propria in CD[22]. Last, CD treatment is based on the use of sulphasalazine and glucocorticoids—two known NF-κB inhibitors[23,24].

Table 3 Allele and genotype frequencies of the three variants associated with CD

Frequencies of the three rare variant alleles

	Number of chromosomes	SNP 8	SNP 12	SNP 13	Total
Unaffected	206	0.04	0.01	0.02	0.07
UC patients	318	0.03	0.00	0.01	0.05
CD patients	938	0.11	0.06	0.12	0.29

Distribution of variant genotypes and associated risks

	Genotype			
	No variant	Simple heterozygous	Homozygous	Compound heterozygous
Distribution				
Unaffected	88	16	0	0
UC patients	145	19	1	0
CD patients	287	133	28	40
Risk for CD				
Relative risk	1	3	38	44
Absolute risk	7×10^{-4}	2×10^{-3}	3×10^{-2}	3×10^{-2}

Genotypes of patients: simple heterozygous, presence of a single rare variant; homozygous, presence of the same variant on both chromosomes 16; compound heterozygous, presence of two different variants; and no variant. Risk of CD for each genotype is computed assuming a prevalence of one per 1,000 and Hardy–Weinberg equilibrium for these markers in the general population. The three rare variants of SNP 8, 12 and 13 were never observed on the same haplotype. UC, ulcerative colitis.

Genetic susceptibility to CD is not limited to chromosome 16 and at least five additional loci have been implicated[25–29]. The recognition of a transduction pathway that, when disregulated, contributes to the pathogenesis of CD will accelerate the discovery of additional susceptibility genes. It will also contribute to the identification of associated environmental factors and focus the search for specific therapies. □

Methods

Families, microsatellite markers and contig construction

A total of 235 CD families (117 simplex nuclear families, 96 multiplex nuclear families, and 22 extended pedigrees, corresponding to a total of 179 CD patients and 251 unaffected relatives) was progressively recruited according to published diagnostic criteria[30]. In addition, 100 multiplex and 59 simplex ulcerative colitis families were recruited from the same hospitals. Written informed consent was obtained from all participants. All relatives from 77 multiplex families were typed for 26 mapped microsatellite markers with an average resolution of 1 cM between SPN and D16S408. We constructed contigs using seven previously localized sequence tag sites (STSs; D16S541, D16S3035, D16S3136, D16S3117, D16S770, D16S416, D16S2623) and subsequently eight additional ones (wi-9288, wi-16305, stgc-17274, stgc-31023, stgc-32374, stSG-30035, wi-5812, D16S766) and 79 new STSs derived from the end sequences of the isolated BAC clones[6].

Clones, sequencing and SNPs

The DNA of BAC clone hb87b10 containing D16S3136 was fragmented by sonication and subcloned in bacteriophage M13. We used sequences from both strands of 706 subclones and from direct primer walking to reconstruct the initial BAC sequence using PhredPhrap (http://www.phrap.org). Identity search in DNA databases identified two overlapping sequenced BACs (AC007334, GenBank; AC007728, GenBank). Homology search, performed on the extended sequence with BLAST v1.4 in GenBank release 114, identified 10 Unigene clusters. The following EST clones corresponding to some of these clusters were obtained from the American Type Culture Collection (http://www.atcc.org) and sequenced completely to identify additional transcribed regions: AI125217, AA417810, AI375427, AA021341, AI090427, AA910520, AA731089. Clones AI090427 and AA910520, corresponding to hs135201, were used to screen a blood leukocyte cDNA library (no. 936202; Stratagene), and retrieved 11 clones of the *IBD1* candidate gene.

A total of 123 STSs, mostly selected from putatively transcribed sequences (EST homologies and GRAIL v2.0 predicted exons), including 11 exons of KIAA0849, was

sequenced following amplifications from the DNA of ten CD patients and two unaffected individuals. Of 35 identified SNPs, SNP 1–11, selected for their rare-allele frequencies greater than 0.06, were typed on 1,272 members of the 235 CD families. SNP 12 and 13 were further identified by sequencing the 11 exons of the candidate *IBD1* gene in 50 CD patients (SNP 1–13, GenBank accession numbers G67943–G67955, are submitted to dbSNP of the National Center for Biotechnology Information) and typed on the same group of individuals. To search for rare variant alleles, we subsequently investigated the 11 exons of 457 CD patients, 159 ulcerative colitis patients and 103 unaffected unrelated individuals. All variant alleles were confirmed by sequencing a second independent amplification product.

Data analysis

Genotypic data were analysed for linkage using the NPL score of GeneHunter v2.0. Data from linkage disequilibrium mapping of CD were analysed initially with the transmission disequilibrium test[7] using a single trio (one affected and both parents) per family. Subsequently, the pedigree disequilibrium test was performed using the PDT 2.11 program[8] to analyse data from all family relatives. We estimated allele frequencies for 3 groups, 418 unrelated CD patients, 159 ulcerative colitis patients and 103 controls (including 78 unaffected, unrelated spouses of CD patients and 25 unrelated CEPH family members).

Received 5 February; accepted 30 March 2001.

1. Calkins, B. M. & Mendelhoff, A. I. in *Inflammatory Bowel Disease* (eds Kirsner, J. B. & Shorter, R. G.) 31–68 (Williams & Wilkins, Baltimore, 1995).
2. Hugot, J. P., Zouali, H., Lesage, S. & Thomas, G. Etiology of the inflammatory bowel diseases. *Int. J. Colorectal Dis.* 14, 2–9 (1999).
3. Hugot, J. P. *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 379, 821–823 (1996).
4. Olavesen, M. G. *et al.* Analysis of single-nucleotide polymorphisms in the interleukin-4 receptor gene for association with inflammatory bowel disease. *Immunogenetics* 51, 1–7 (2000).
5. Hugot, J. P. *et al.* Mutation screening in the CD19 and CD43 (sialophorin) genes in Crohn Disease patients. *Gastroenterology* 116, A740 (1999).
6. Zouali, H. *et al.* Refined mapping of the inflammatory bowel disease 1 gene. *Eur. J. Hum. Genet.* (submitted).
7. Spielman, R. S., McGinnis, R. E. & Ewens, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52, 506–516 (1993).
8. Martin, E. R., Monks, S. A., Warren, L. L. & Kaplan, N. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* 67, 146–154 (2000).
9. Kuster, W., Pascoe, L., Purrmann, J., Funk, S. & Majewski, F. The genetics of Crohn's disease: complex segregation analysis of a family study with 265 patients with Crohn's disease and 5387 relatives. *Am. J. Med. Genet.* 32, 105–108 (1989).
10. Monsen, U., Iselius, L. & Hellers, G. Evidence for a recessive gene in Crohn's disease. *Acta Chir. Scand.* 559 (suppl.), 7–42 (1990).
11. Orholm, M. *et al.* Investigation of inheritance of chronic inflammatory bowel disease by complex segregation analysis. *Br. Med. J.* 306, 20–24 (1993).
12. Colombel, J. F. *et al.* Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* 111, 604–607 (1996).
13. The IBD International Genetics Consortium. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *Am. J. Hum. Genet.* 68, 1165–1171 (2001).
14. Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kB. *J. Biol. Chem.* 276, 4812–4818 (2001).
15. Inohara, N., Ogura, Y., Chen, F. F., Muto, A. & Nuñez, G. Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *J. Biol. Chem.* 276, 2551–2554 (2001).
16. Inohara, N. *et al.* Nod 1, an Apaf-1-like activator of caspase-9 and nuclear factor kB. *J. Biol. Chem.* 274, 14560–14567 (1999).
17. Bertin, J. *et al.* Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-kB. *J. Biol. Chem.* 274, 12955–12958 (1999).
18. Inohara, N. *et al.* An induced proximity model for NF-kB activation in the Nod1/RICK and RIP signaling pathways. *J. Biol. Chem.* 275, 27823–27831 (2000).
19. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 282, 2085–2088 (1998).
20. Sundberg, J. P., Elson, C. O., Bedigian, H. & Birkenmeier, E. H. Spontaneous, heritable colitis in a new substrain of C3H/HeJ mice. *Gastroenterology* 107, 1726–1735 (1994).
21. McKay, D. M. Intestinal inflammation and the gut microflora. *Can. J. Gastroenterol.* 13, 509–516 (1999).
22. Schreiber, S., Nikolaus, S. & Hampe, J. Activation of nuclear factor kB in inflammatory bowel disease. *Gut* 42, 477–484 (1998).
23. Auphan, N., DiDonato, J. A., Rosette, C., Helmberg, A. & Karin, M. Immunosuppression by glucocorticoids: inhibition of NF-kB activity through induction of IkB synthesis. *Science* 270, 286–290 (1995).
24. Wahl, C., Liptay, S., Adler, G. & Schmid, R. M. Sulfasalazine: a potent and specific inhibitor of nuclear factor kB. *J. Clin. Invest.* 101, 1163–1174 (1998).
25. Satsangi, J. *et al.* Two stage genome wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosome 3, 7 and 12. *Nature Genet.* 14, 199–202 (1996).
26. Hampe, J. *et al.* Linkage of inflammatory bowel disease to Human chromosome 6p. *Am. J. Hum. Genet.* 65, 1647–1655 (1999).
27. Cho, J. H. *et al.* Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Hum. Mol. Genet.* 9, 1425–1432 (2000).
28. Duerr, R. H., Barmada, M. M., Zhang, L., Pfutzer, R. & Weeks, D. E. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am. J. Hum. Genet.* 66, 1857–1862 (2000).
29. Rioux, J. D. *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am. J. Hum. Genet.* 66, 1863–1870 (2000).
30. Lennard-Jones, J. E. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol.* 170 (suppl.), 2–6 (1989).

Acknowledgements

We acknowledge patients and their families, and thank family recruitment doctors: J. Belaiche, B. Bonaz, Y. Bouhnik, G. Cadiot, S. Cucchiara, B. Cessaris, J. J. Delchier, B. Duclos, J. L. Dupas, J. P. Galmiche, J. P. Gendre, D. Golfain, C. Grimaud, D. Heresbach, A. Lachaux, H. Lamraite, C. Lenaerts, E. Lerebours, V. Levy, R. Löfberg, H. Malchow, P. Marteau, A. Morali, F. Pallone, S. Pena, A. Rotenberg, L. Rousseau, J. Schmitz, F. Shanahan, I. Sobhani, H. Svensson, A. Van Gossum, M. Van Winckel and M. Veyrac. For assistance we are grateful to J. C. Beaudoin, F. Chareyre, C. Giudicelli, T. Hung Bui, M. Legrand, A. Marcadet, A. Martins, C. de Toma, E. Tubacher. We thank H. Cann for critically reading the manuscript. This project received support from European Community, MENRT, INSERM, Direction Générale de la Santé, Association François Aupetit, IRMAD and the Swedish Society of Medicine.

Correspondence and requests for materials should be addressed to G.T. (e-mail: thomas@cephb.fr).

A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease

Yasunori Ogura*†, Denise K. Bonen‡†, Naohiro Inohara*, Dan L. Nicolae§, Felicia F. Chen*, Richard Ramos‡, Heidi Britton‡, Thomas Moran‡, Reda Karaliuskas‡, Richard H. Duerr‖, Jean-Paul Achkar¶, Steven R. Brant#, Theodore M. Bayless#, Barbara S. Kirschner☆, Stephen B. Hanauer‡, Gabriel Nuñez*†† & Judy H. Cho‡††

* *Department of Pathology and Comprehensive Cancer Center, The University of Michigan Medical School, Ann Arbor, Michigan 48109, USA*
‡ *The Martin Boyer Laboratories, Gastroenterology Section, Department of Medicine, The University of Chicago Hospitals, Chicago, Illinois 60637, USA*
§ *Department of Statistics, and* ☆ *Department of Pediatrics, University of Chicago, Chicago, Illinois 60637, USA*
‖ *Department of Medicine and Center for Genomic Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
¶ *Department of Gastroenterology, The Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA*
\# *The Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA*
† *These authors contributed equally to this work*
†† *These authors share senior authorship*

Crohn's disease is a chronic inflammatory disorder of the gastro-intestinal tract, which is thought to result from the effect of environmental factors in a genetically predisposed host. A gene location in the pericentromeric region of chromosome 16, *IBD1*, that contributes to susceptibility to Crohn's disease has been established through multiple linkage studies[1–6], but the specific gene(s) has not been identified. *NOD2*, a gene that encodes a protein with homology to plant disease resistance gene products is located in the peak region of linkage on chromosome 16 (ref. 7). Here we show, by using the transmission disequilibium test and case-control analysis, that a frameshift mutation caused by a cytosine insertion, 3020insC, which is expected to encode a truncated NOD2 protein, is associated with Crohn's disease. Wild-type NOD2 activates nuclear factor NF-κB, making it responsive to bacterial lipopolysaccharides; however, this induction was deficient in mutant NOD2. These results implicate *NOD2* in susceptibility to Crohn's disease, and suggest a link between an